

Open Research Online

The Open University's repository of research publications and other research outputs

Understanding student experience: a pathways model

Conference or Workshop Item

How to cite:

Edwards, Chris and Gaved, Mark (2020). Understanding student experience: a pathways model. In: L@S '20: Proceedings of the Seventh ACM Conference on Learning @ Scale, 12-14 Aug 2020, Virtual event, USA, pp. 265–268.

For guidance on citations see [FAQs](#).

© 2020 The Authors



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1145/3386527.3406724>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Understanding student experience: a pathways model

Chris Edwards
Open University
Milton Keynes, UK
chris.edwards@open.ac.uk

Mark Gaved
Open University
Milton Keynes, UK
mark.gaved@open.ac.uk

ABSTRACT

As universities increasingly teach at scale, new challenges are introduced and compounded where students are offered greater choice. A key challenge is to maintain an understanding of the student experience within the huge increase in variations in student study path. This understanding is necessary to provide feedback to both faculty and students, and institutionally for the enhancement of quality. This is the first description of one fresh approach to this challenge. Whilst based on the experience within a large distance learning university, the findings are relevant to all institutions working at scale. Moving from a traditional relational structure to a multi-model database makes it possible to quickly design study path queries to explore the richness of available data. We provide an overview of this approach that could be applied by other universities and higher education institutions where data is not being fully utilised.

Author Keywords

Higher Education; Course Sequences; Academic Pathways; Graph Visualization; Multi-model Database; Student Experience; Quality Enhancement.

CSS Concepts

•Computing methodologies~Modeling and simulation~Model development and analysis~Modeling methodologies

INTRODUCTION

As universities increasingly teach at scale new challenges are introduced. These are further compounded as institutions increasingly offer greater levels of choice to students (in terms of which modules to study, which order to study them, and how long to extend study before qualification). A key challenge is to maintain an understanding of the student experience and the results from the huge increase in variations in student study paths: particularly because of the fragmentation of module cohorts into sub-cohorts [2]. This understanding is necessary to provide feedback to both students and faculty and is required institutionally to enable the continued enhancement of quality [5]. This paper is the first description of one fresh approach to this challenge. Whilst based on the experience within a large distance learning university, the findings are relevant to all institutions working at scale. Moving from a traditional relational structure to a multi-model database makes it possible to quickly design study path queries to explore the richness of available data. We provide an overview of this approach that could be applied by other universities and higher education institutions where data is not being fully utilised.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
L@S '20, August 12–14, 2020, Virtual Event, USA
© 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-7951-9/20/05...\$15.00
<https://doi.org/10.1145/3386527.3406724>

first description of one fresh approach to this challenge. Whilst the approach is based on the experience within a large distance learning university, the findings are relevant to all institutions working at scale. It demonstrates that a new approach to data is key to facilitating the analysis of student study pathways. For many years, this University has offered great flexibility of study and as wide a study choice as it is possible to offer through a modular approach. By design, the University holds high levels of data for all student study. However, whilst it is possible to create bespoke queries of this data, experience has shown this to be too resource-intensive to readily enable different analyses of the student experience; leaving colleagues frustrated that many of their questions about the experience of students within their qualifications remained unanswered. By moving from a traditional relational database structure to a graph database, a powerful graphical query language makes it possible to quickly design study path queries that more readily explore the richness of available data. In this paper, we provide an overview of this approach that could be applied by other universities and higher education institutions where data is not been fully utilised to enhance the student experience, and describe the next steps.

Experience in one university, from years of working with data arranged conventionally in a relational database structure, had shown that whilst it was always possible to answer specific study path questions, the level of resource required often made the question practically unanswerable. The queries involved were usually complex and required specialist knowledge. By 2017, database technology had developed to include successful graph and multi-model solutions and so we decided to approach the question afresh. While there was no shortage of data, the struggle was to derive meaning around the student experience.

The new database technologies much more readily supported linked data, through graph methods. Multi-model databases support different database structures (including graph, document and key-value) within a single database. Therefore, how would we organize the data if we were starting from now and intending to make the student and their experience central to how we worked with and developed curriculum? How simple could the data structure be to aid analysis and understanding?

One intended outcome is to create a model that makes data more accessible to the people that need to learn from it. The ideas expressed by Prestigiacomo et al around the concept of translucence are of value [9]. Namely the linked values of

visibility, awareness and accountability. For us, the first step is to make students' study choices and experience visible. Through analysis we can better understand and therefore gain and share awareness, which in turn can support proper accountability, i.e. informed actions.

Terminology

As confusions can be introduced through differing uses of terms, this is how we will use some key terms within this paper.

Course: the programme of study leading to a qualification.

Key Introductory module – A module designed as the entry point to degree study

Module: a coherently designed and delivered unit of teaching and learning on a particular subject. This will usually also be smallest portion of learning that can be awarded credit towards a qualification

Presentation: due in main to the logistics of managing teaching resource and ensuring quality of assessment, most universities will offer each module with discreet start, end, and assessment dates. Each time a module is offered, each presentation, it maybe with different staff teaching, and variations in content and assessment.

Programme of study: the modules that need to be passed before a qualification can be awarded. Each qualification will have its own rules about compulsory and optional modules, and the order of study. There may be institutional rules also applying. The programme of study for a qualification will include all the potential modules and rules for that qualification. For an individual, their programme of study comprises their particular choices, both planned and taken towards their intended qualification.

Qualification: a widely recognised certificate a university awards to students on successful completion of a particular programme of study.

Smallest creditable element (SCE) – this element is the combination of student-module-presentation, nothing smaller than this is awarded credit. We can also think of this as a student study attempt on a module. The smallest creditable element may differ from university to university.

Study attempt – when a student begins to study a module and passes a particular point (determined by each university) to avoid the very early dropouts that would otherwise skew the data, they make a *study attempt*, irrespective of the outcome of this attempt. For the model described here, we generally grouped outcomes into: pass, fail, defer, withdraw. More detail was available within the database to be called on if necessary.

Study path/route: This is the chronological series of modules actually attempted by an individual student. It includes modules failed as well as passed, and those from which a student has withdrawn: perhaps to return to later (deferred).

METHODOLOGY

In order to represent the student study data within a graph structure, an initial step was to define which data are initially aligned with vertices and which to the edges connecting them. The simplest arrangement is to define the vertices as representing the SCE, containing the student experience of their attempt to study a module. The edges would then represent the link between starting one module and starting the next and record the time gap between these. The edges have direction, showing the order in which modules were studied. In the database we have named this vertex collection as 'Study' and the edge collection as 'Path'. There is an analogous pair of collections for the curriculum. These are called 'Module' and 'Qualification'.

This group of four collections is sufficient to allow every student pathway study experience to be explored and to hold the record of every programme of study within the University. They structure the data within the database to model the relationships between them from the perspective of the student. It is important to note that no new data were sought. This model contains existing data, organized in a new way [4]. The values stored in the vertices collection were taken from the University's existing data warehouse. The edges for the paths collection were calculated programmatically from the existing data, with scripts written in Python.

There were several potential databases capable of enabling this metaphor to be realised. We selected ArangoDB for a several reasons, including: that it was both free and open source; had the potential to be fully scalable to meet institutional needs; offered Azure and AWS integration; uses a single, elegant query language; is based on sound technology; has complete flexibility.

In order to use the model, queries are required. To make it as accessible as possible, these queries must be readily available and straightforward to modify for a specific context. A query library was therefore established as a documented repository. An example of a very simply query is one that will plot the study path of a specific student.

```
LET student = "the-id-for-one-student"
For      v,e,p      IN      1..40      OUTBOUND
CONCAT("Study/",student,"-start-start") Path
RETURN p
```

The database recognises the linked relationships and by default this query returns the data plotted as a graph, like the one on the left in Figure 1. The plot on the right is the same base query adapted for the Module and Qualification to return the programme of study for a single qualification

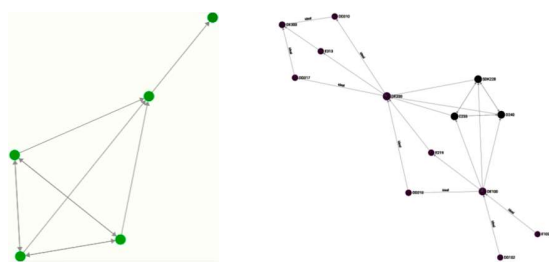


Figure 1. Left: The study path of a single student, with the 'start' vertex removed for clarity. Note: the group of three vertices at the bottom left are connected with double ended arrows, showing they were studied together; Right: the plot of a programme of study for a single qualification.

Over three years we have worked with faculty academics who have outstanding pathway questions, exploring the model's capacity to answer these. These include an exploration of onward study paths, where colleagues in the faculty of Science, Technology, Engineering and Mathematics (STEM) had been trying to establish the study routes students take through their Science curriculum. As a first step, we considered the cohort of students who began their study with the University in one specific presentation of one Science Key Introductory module and collate their ongoing study.

Another case study involved supporting those leading the Psychology qualifications in reviewing their effectiveness. For this study, all students who had study paths consistent with the qualifications were included and the relative effectiveness of each of these study paths compared. Two further case studies looked at study choices in different subjects. One in Computing and IT and the other in Language Studies.

RESULTS

The results revealed a greater than anticipated complexity. The left-hand plot in Figure 2 illustrates this. It shows the aggregated onward study of a single cohort of students ($n=937$) from one Science Key Introductory module [3]. We have repeatedly found that a significant proportion of study paths are travelled by very few students and a small number of study paths carry the majority of students. The pattern has a recognisable 'long tail'. The plot on the right-hand side of Figure 2 demonstrates this further, as it only includes module-module transitions chosen by at least ten students. The plot reduces to one that is possible to interpret visually.

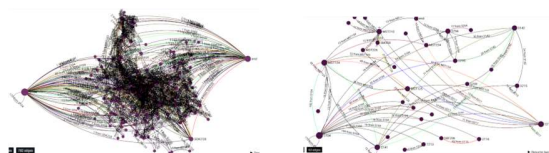


Figure 2. Two views of the onward study of a cohort of students taking one presentation of a Key Introductory Science module. Left: the full but incomprehensibly busy picture. Right: module to module transitions chosen by at least 10 students. The cohort contained 937 students.

For the first year of the degree in Language Studies, we found 1436 potential study routes and instead of plotting the results of the query, for this case study we introduced a textual notation for the study paths. This enabled the aggregated study paths to be exported to familiar software, in this case Microsoft Excel, for further analysis. As an example, the study path in the left-hand side of Figure 1 could be, (Ma-P1, Mb-P1, Mc-P1)->Md-P2->Me-P3: where Ma is Module A and P1 is presentation 1. Concurrent study is grouped alphabetically within brackets.

The case study in which we considered a review of a whole qualification, in which there are 64 study paths designed into the programme of study revealed several valuable insights that would not emerge in the conventional processes around reporting. Two of these are: that the positive impact of choosing to study one module over another in the first year of study continued for at least two further modules; that although numbers on all presentations through the qualification were very healthy, no student had managed to successfully attempt the whole programme of study due to timing, not failed attempts – even though technically feasible.

The model readily enables other analyses. For example, any university where there are options to extend study over a longer period, will have an interest on the impact of study gap/overlap on study outcome and the student experience [7,8]. It was possible to demonstrate this analysis for a portion of one programme of study. The findings of this limited study indicate a clear optimum rate of study and will lead to further work on this question.

DISCUSSION

Invariably, when we discussed the pathways model with colleagues, their responses fell into two main groups: those focused on improving curriculum, and those wanting to provide students with advice and guidance. From experience of applying the model over three years, it is clear there is potential to support both. The results highlight the importance of not simply focusing on the major study paths as a long tail of a significant proportion of students take less travelled study routes.

The multi-model database approach made it possible for the data model to be designed as a metaphor for the student experience. This makes it instantly more accessible to a wider proportion of users and for analysis [1]. This assists with the cultural change required to move from the current structures for reviewing curriculum developed over decades. It also aids those who are uncomfortable, intimidated or confused when faced with current dashboards. The database can present data to AI platforms and the ArangoDB query language includes machine learning queries. We have yet to test both of these.

An obvious advantage of the database's native representation of linked data as a graph is that it is a powerful way to reveal patterns, both expected and unexpected. For example, the

plot of a student study path in Figure 1, immediately shows that this student studied their first three modules concurrently.

The current version of the pathways model does not yet include the fine-grained data that does exist for many activities and interactions within module presentations. This was due to a deliberate decision to start as simply as possible.

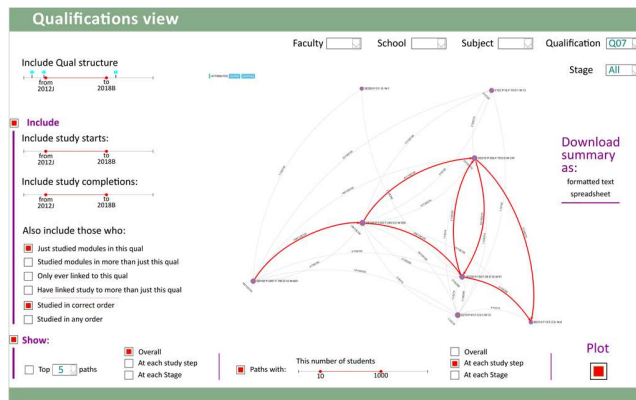


Figure 3. A mock-up of a qualification based interactive visual query of the model.

CONCLUSION

Through a reorganization of existing data, we have built a straightforward data model as a metaphor for the student study experience. By implementing this model within a multi-model database, it can be used to readily answer many questions that have, in some cases, remained intractable for years. The database recognises relationships in query results and natively produces graphical arrangements of these. The benefits of this are to make the data more accessible to a wider range of those who need to learn from it, including staff and students, and to further analysis using conventional and deep learning methods. The model, technology and approach are of value to any university grappling with student experience.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of colleagues who have given their time and expertise in working with us to trial this pathways model.

REFERENCES

- [1] Steven Alter. 2017. Nothing is more practical than a good conceptual artifact... which may be a theory, framework, model, metaphor, paradigm or perhaps some other abstraction. *Inf. Syst. J.* 27, 5 (2017), 671–693. DOI:https://doi.org/10.1111/isj.12116
- [2] Geoffrey Angus, Richard Diehl Martinez, Mitchell L Stevens, and Andreas Paepcke. 2019. Via: Illuminating Academic Pathways at Scale. In *L@S '19: Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*, Association for Computing Machinery. New York, NY, United States, Chicago IL USA. DOI:https://doi.org/10.1145/3330430.3333623
- [3] Clow, Doug; Coughlan, Tim; Cross, Simon; Edwards, Chris; Gaved, Mark; Herodotou, Christothea; Nguyen, Quan; Rienties, Bart; Thorne, Sam and Ullmann, Thomas (2019). *Scholarly insight Winter 2019: a Data wrangler perspective*. Open University, Milton Keynes. Available at, <http://oro.open.ac.uk/59646/>
- [4] Edwards, Chris (2017). Understanding student experience in the age of personalised study. In: *ALT Online Winter Conference 2017*, 12-13 Dec 2017. Available at <http://oro.open.ac.uk/53745/>
- [5] Asmaa Elbadrawy and George Karypis. 2019. UPM: Discovering Course Enrollment Sequences Associated with Success. In *LAK19: Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, Association for Computing Machinery, New York, NY, United States, Tempe, AZ, USA. DOI:https://doi.org/10.1145/3303772.3303799
- [6] Christopher Krauss, Agathe Merceron, and Stefan Arbanowski. 2019. The Timeliness Deviation: A novel Approach to Evaluate Educational Recommender Systems for Closed-Courses. In *LAK19: Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, Association for Computing Machinery, New York, NY, United States, Tempe, AZ, USA. DOI:https://doi.org/10.1145/3303772.3303774
- [7] Sara Morsy and George Karypis. 2019. A Study on Curriculum Planning and Its Relationship with Graduation GPA and Time To Degree. In *LAK19: Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, Association for Computing Machinery, New York, NY, United States, Tempe, AZ, USA. DOI:https://doi.org/10.1145/3303772.3303783
- [8] Zachary A Pardos, Hung Chau, and Haocheng Zhao. 2019. Data-Assistive Course-to-Course Articulation Using Machine Translation. In *L@S '19: Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*, Association for Computing Machinery. New York, NY, United States, Chicago IL USA. DOI:https://doi.org/10.1145/3330430.3333622
- [9] Rita Prestigiacomo, Roger Hadgraft, Jane Hunter, Roberto Martinez-Maldonado, Lori Locker, Simon Knight, and Elise van den Hoven. 2020. Learning-centred Translucence: An Approach to Understand How Teachers Talk About Classroom Data Human-centred design; evidence-based decision-making ACM Reference format. In *LAK '20: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, Association for Computing Machinery, New York, NY, United States, Frankfurt, Germany. DOI:https://doi.org/10.1145/3375462.3375475